

Chapter 9

Complexity, Emergence and Molecular Diversity via Information Theory

Francisco Torrens

Institut Universitari de Ciència Molecular, Universitat de València, Spain

Gloria Castellano

Universidad Católica de Valencia San Vicente Mártir, Spain

ABSTRACT

*Numerous definitions for complexity have been proposed with little consensus. The definition here is related to Kolmogorov complexity and Shannon entropy measures. However, the price is to introduce context dependence into the definition of complexity. Such context dependence is an inherent property of complexity. Scientists are uncomfortable with such context dependence that smacks of subjectivity, which is the reason why little agreement is found on the meaning of the terms. In an article published in *Molecules*, Lin presented a novel approach for assessing molecular diversity based on Shannon information theory. A set of compounds is viewed as a static collection of microstates that can register information about their environment. The method is characterized by a strong tendency to oversample remote areas of the feature space and produce unbalanced designs. This chapter demonstrates the limitation with some simple examples and provides a rationale for the failure to produce results that are consistent.*

INTRODUCTION

Complexity is conceptually impressive; it is a discipline that can be applied to energy packets, traffic, neurones, bourse markets, molecules inside the cell, *etc.*, to any system formed by elements that interact between them in an apparently random

way and then, without understanding why and how, something happens (Mitchell, 2009). The emergent properties are impressive; they leave us astonished and state two basic questions. (1) Is there a science that could arrive to describe in a satisfactory way the internal laws of complexity? (2) This new that appears, was it there before and is it a question of being more refined with the calculations or is it really unforeseen?

DOI: 10.4018/978-1-4666-2077-3.ch009

We shall give an example from chemistry (Pullman, 1994). Imagine that one knows everything on the structure of the water molecule. Can this person foresee the transition from liquid to ice? Or does it happen at a different level? We think that it is fundamentally different because the second law of thermodynamics, *e.g.*, appears when there are more than one particle not only one. Physicists are not used to like the statement because they think that if they go shaking out the system top-down, they could become to understand all its features. They are used to be more reductionist but we as chemists have a different approach. We at the laboratory build up new complex systems instead of deconstructing the ones existing already. Moreover we study them in a more holistic way.

In earlier reports the fractal hybrid-orbital analyses of the protein tertiary structure were carried out (Torrens, 2000, 2001). Valence topological charge-transfer indices for molecular dipole moments were obtained (Torrens, 2004). Information-entropy molecular classification was applied to local anaesthetics (Castellano & Torrens, 2009; Torrens & Castellano, 2006) and inhibitors of human immunodeficiency virus type 1 (Torrens & Castellano, 2009, 2010, in press, a). It was reported the structural classification of complex molecules by artificial intelligence techniques (Torrens & Castellano, 2012). It was published the structural classification of complex molecules by information entropy and equipartition conjecture (Torrens & Castellano, in press, b). It was performed the molecular diversity studies of the bond-based linear indices of the non-stochastic and stochastic edge adjacency matrix of the physicochemical properties of organic molecules (Marrero-Ponce *et al.*, 2010) and novel coumarin-based tyrosinase inhibitors discovered by the Organisation for Economic Co-operation and Development (OECD) principles-validated quantitative structure-activity relationship (QSAR) approach from an enlarged balanced database (Le-Thi-Thu *et al.*, 2011). The following subsections describe the problem of complexity,

the concept of emergence, the entropy as a case study of emergence and the complexity in the group method of data handling (GMDH)-type neural networks. Then the computational method is explained. In next sections the calculation results are presented and discussed. The final section summarizes the conclusions.

THE PROBLEM OF COMPLEXITY

The study of *complex systems* emerged as a field although a good definition eluded formulation (Standish, 2001). Attempts go back to Shannon's inception of *information theory* (Shannon, 1948ab). Edmonds (1999) provided a good historical survey of *complexity*. It is important the concept of *Kolmogorov algorithmic information complexity* (KC) introduced by Kolmogorov (1965), Chaitin (1966) and Solomonoff (1964ab). Given a particular universal Turing machine (UTM) U , KC of a string of characters is the length of the shortest program running on U that generates the description. It has two main problems. (1) The dependence on U as there is no unique way of specifying it. The invariance theorem guarantees that any two UTMs, U and V , will agree on the complexity of a string x up to a constant independent of x (Li & Vitányi, 1997). However, for any two descriptions x and y there will be two machines U and V disagreeing on whether x is more complex than y or vice versa. (2) Random sequences have maximum complexity, as by definition a random sequence can have no generating algorithm shorter than simply listing the sequence, which as Gell-Mann (1994) pointed out contradicts the notion that random sequences should contain no information. The first problem of what reference machine to choose is a symptom of context dependence of complexity. Given a description x any value of complexity can be chosen for it by choosing an appropriate reference machine. It would seem that complexity is in the eye of the beholder (Emmeche, 1996). Is com-

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/complexity-emergence-molecular-diversity-via/69462

Related Content

Activity Theory

Lars Taxén (2010). *Using Activity Domain Theory for Managing Complex Systems* (pp. 65-77).

www.irma-international.org/chapter/activity-theory/39672

The Role Micro-Blogging Plays in Informal Communication and Knowledge Sharing Activities Within Universities: A Review of The Literature

Shakiba Kazemian (2018). *International Journal of Knowledge and Systems Science* (pp. 18-31).

www.irma-international.org/article/the-role-micro-blogging-plays-in-informal-communication-and-knowledge-sharing-activities-within-universities/224957

A Self-Organizing Neural Network to Approach Novelty Detection

Marcelo Keese Albertiniand Rodrigo Fernandes de Mello (2010). *Intelligent Systems for Automated Learning and Adaptation: Emerging Trends and Applications* (pp. 49-71).

www.irma-international.org/chapter/self-organizing-neural-network-approach/38450

Ontology-Based Competence Management in Pharmacy

J. Puustjärviand L. Puustjärvi (2013). *International Journal of Adaptive, Resilient and Autonomic Systems* (pp. 94-112).

www.irma-international.org/article/ontology-based-competence-management-pharmacy/77634

A Triple Cornerstone Framework for Software Startups: A Systems Thinking-Based Analysis

Narendranath Shanbhagand Eric Pardede (2021). *Handbook of Research on Modeling, Analysis, and Control of Complex Systems* (pp. 60-90).

www.irma-international.org/chapter/a-triple-cornerstone-framework-for-software-startups/271034